Ruoheng Du & Victoria Fu
Professor Sebastiano Manzan
ECON-UB 251 Econometrics I
May 15th, 2023

Explaining and Predicting ARKK ETF with Fama-French Four-Factor Model and Black-Litterman Model

I. Introduction

The stock market is a dynamic and complicated system that can be difficult to forecast and several asset pricing models have been developed to assist investors in improving understanding. This project focuses on the actively managed Exchange Traded Fund (ETF) ARKK, which pursues long-term capital gain by investing in companies with disruptive innovation that are predicted to experience hyper-growth in the future, and aims to use the Fama-French four-factor model and the Black-Litterman model to explain and predict its stock returns.

While numerous scholars have investigated the relationship between stock returns and other characteristics such as industry-specific factors and sentiment-related views, little study has been done especially on hyper-growth technology businesses such as those in which ARKK invests. This study is important because it explains how ARKK's unique investing strategy and assets affect its performance. Furthermore, considering ARKK's focus on disruptive innovation, our findings may have broader implications for investors interested in hyper-growth businesses.

Our analysis suggests that both the Fama-French four-factor model and the Black-Litterman model are extremely essential in explaining ARKK's stock returns, while the Black-Litterman model with lagged ARKK excess return performs better. This showcases that the monthly returns of ARKK is closely related to its characteristics of active management and hyper-growth under the current macroeconomic environment.

Our findings are relevant for ARKK investors because they imply that these models can be used to make informed investing decisions. Furthermore, these models can help investors understand the fundamental factors that drive ARKK's returns, allowing for more accurate estimates of future performance.

II. Literature Review

In the field of finance, the capital asset pricing model has been a popular study topic. Markowitz first proposed portfolio theory as a way to balance risk and reward (6). In the 1960s, Sharpe developed the capital asset pricing model, which utilizes market return as the sole variable for explaining returns (428).

However, Fama and French discovered that the beta value of the CAPM could not entirely explain the difference in excess returns observed in the market. As a result, they created the Fama-French three-factor model, a new three-factor model that classified the primary components into three categories: market factor, size factor, and value factor (425-428). For instance, Wen and Yin utilized the Fama-French three-factor model in their empirical analysis of 80 pharmaceutical and biological industry listed companies, and concluded that the market premium factor, size factor, and book-to-market ratio factor adequately explain the excess returns of these companies (87).

Later, some researchers, such as Carhart, discovered that the Fama-French three-factor model had limits in explaining differences in excess returns (59-60). As a result, Carhart modified the model by including a momentum factor, resulting in the Fama-French four-factor model (i.e., Carhart four-factor model). This has been proven in the Moroccan market. Although both models have a limited ability to forecast cross-sectional returns on the Casablanca Stock Exchange (CSE), the Fama-French four-factor model has slightly greater explanatory power than the Fama-French three-factor model (20).

To overcome the constraints of the traditional mean-variance method, Black and Litterman suggested a model that starts with an equilibrium expected return in which demand equals asset supply and then combines market equilibrium with investor perspectives (He and Litterman 8). In the context of twenty largest capitalization stocks listed on NASDAQ and NYSE, Li et al. (515) found that the Black-Litterman model is superior to the single index model, Fama-French three-factor model, Fama-French four-factor model, and can generate relatively high returns and relatively low risk. A recent research on sentiment-related factors (Ji 4) in the chinese blockchain industry has found that investor sentiment has an impact on stock

returns, and adding sentiment as a new independent variable outside the Fama-French three-factor Model can effectively enhance the explanatory power of the new model.

III. Model and Data

3.1 ARKK ETF

ARKK is an actively managed ETF and it aims to achieve long-term capital growth. The fund invests in a range of US and non-US stocks that are related to the theme of disruptive innovation, which refers to products or services that can potentially revolutionize the way things are done in the world. The fund seeks to invest in hyper-growth stocks that are expected to deliver substantial profits in the distant future. While Facebook and Amazon are often considered growth stocks, ARKK invests in companies that exhibit even higher returns and volatilities. For example, Tesla had a return of 7 times its initial investment in 2020, and Shopify saw a compounded return of approximately 1.9 times between 2017 and 2020. As of December 31, 2022, the ETF has a 37.9% weight in Information Technology and 33.4% in Health Care industries. Their top 3 holdings of this ETF are Zoom Video Communications for 9.6%, Exact Sciences for 8.6%, and Tesla for 7.0%.

3.2 Regression Model

We will do two regression models based on the Fama-French four-factor model and the Black-Litterman model to see which model can best predict the return of ARKK. The Fama-French four-factor model is one of the most widely used asset pricing models that uses four factors to explain stock returns: market risk, size, value, and momentum. It is a more complete approach to asset pricing than the CAPM as it incorporates additional elements that have been demonstrated to be critical predictors of stock returns. The Black-Litterman model is also a portfolio optimization approach that attempts to add investor's subject views into the CAPM. It is created as an enhancement on the classic Markowitz portfolio optimization model. It offers a more intuitive and flexible asset allocation strategy than standard models and is frequently used in investment management.

The data of ARKK monthly return used in the regression are collected from Yahoo

Finance and the data of market return, SMB, HML, MOM and risk-free rate are all collected from Kenneth R. French's Data Library.

Dependent Variable:

- ARKK monthly excess return (%): This is calculated by the monthly return of ARKK minus the monthly risk-free rate.

Independent Variables:

- Excess Market Return (%): Calculated by monthly market return minus the monthly risk-free rate.
- SMB (%): One of the three factors of the Fama-French model, which accounts for the publicly traded companies with small market capitalizations that generate higher returns (Fama and French). It is the difference in returns between small-cap and large-cap stocks.
- HML(%): One of the three factors of Fama-French model, which accounts for value stocks with high book-to-market ratios that generate higher returns in comparison to the market (Fama and French) and it captures value equities' proclivity to outperform growth stocks. It is the difference in returns between high book-to-market (value) and low book-to-market (growth) companies.
- MOM (%): The monthly momentum factor is one of the four factors of the Fama-French four-factor model, and it is calculated by the equal weighted average of the highest performing firms during a one-month period subtracting the equal weighted average of the lowest performing firms. MOM factor indicates the proclivity of stocks that are well performed in the past to continue doing well in the future, as well as stocks that are poorly performed in the past to continue performing poorly in the future.
- Lagged ARKK monthly excess return (%): We will take the lagged one-month excess return of ARKK as our view in the Black-Litterman model and this represents the return earned by ARKK in the previous one month. We take the lagged excess ARKK as our view because we think time series will influence the return of ARKK and it will also influence the Fama-French four-factor model, so, we take this as our view to tackle the issue of autocorrelation.

In the first regression, which is based on the Fama-French four-factor model, we will

include excess market return, SMB, HML and MOM as independent variables while ARKK excess return as the dependent variable, and the formula in this case is shown as below:

$$R_{ARKK} - R_f = \alpha + \beta_1(R_{Market} - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 MOM + \varepsilon .$$

In the second regression based on the Black-Litterman model, we will use excess market return, SMB, HML, and lagged excess ARKK return as the independent variables, and the formula is shown as below:

$$R_{ARKK} - R_f = \alpha + \beta_1(R_{Market} - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 Lagged(R_{ARKK} - R_f) + \varepsilon.$$

IV. Result

4.1 The Relative Performance of ARKK

We chose the NASDAQ composite index, which mostly includes tech stocks related to the theme of innovation and is representative market index for the technology industry which ARKK focuses on, as a benchmark to further investigate ARKK's relative performance. The table below contains annualized excess returns of ARKK and NASDAQ composite index, which is calculated by averaging the monthly returns and multiplying them by 12. From the table below, during 2018 and 2019, ARKK slightly outperformed the benchmark. However, in 2020, when U.S. tech stocks benefited from work and study from home amid COVID-19, ARKK achieved an impressive return of 57.03%, which greatly outperformed the benchmark. In 2021 and 2022, on the contrary, the ARKK greatly underperformed the benchmark because the expectations of Fed rate hikes would increase the risk-free rate and lowered the valuations of growth stocks, especially the hyper-growth stocks.

	2018	**2019**	**2020**	**2021**	**2022**
ARKK return (%)	-174.05	-174.28	57.03	-22.41	-240.23
Benchmark Return (%)	-182.39	-181.30	-5.91	18.18	-178.28
ARKK-Benchmark Return (%)	8.34	7.02	62.94	-40.59	-61.95

4.2 ARKK under the Fama-French Four-Factor Model

	Dependent variable:						
	ARKK.RF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mkt.RF	0.019***		0.017***	0.019***	0.017***	0.017***	0.017***
	(0.003)		(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
SMB		0.028***	0.018***		0.017***	0.017***	0.017***
		(0.006)	(0.005)		(0.005)	(0.005)	(0.005)
HML				-0.006*	-0.005*		-0.005*
				(0.003)	(0.003)		(0.003)
MOM						-0.004	-0.004
						(0.006)	(0.006)
Constant	-0.107***	-0.090***	-0.103***	-0.107***	-0.104***	-0.091***	-0.091***
	(0.015)	(0.017)	(0.013)	(0.014)	(0.013)	(0.023)	(0.022)
Observations	60	60	60	60	60	60	60
R^2	0.488	0.245	0.580	0.517	0.603	0.583	0.607
Adjusted R^2	0.479	0.232	0.565	0.500	0.582	0.561	0.578
Akaike Inf. Crit.	-89.062	-65.802	-98.961	-90.629	-100.352	-97.466	-98.929
Bayesian Inf. Crit.	-82.779	-59.519	-90.583	-82.251	-89.880	-86.994	-86.363
Residual Std. Error	0.111 (df = 58)	0.135 (df = 58)	0.102 (df = 57)	0.109 (df = 57)	0.100 (df = 56)	0.102 (df = 56)	0.100 (df = 55)
F Statistic	55.194*** (df = 1; 58)	18.818*** (df = 1; 58)	39.321*** (df = 2; 57)	30.528*** (df = 2; 57)	28.337*** (df = 3; 56)	26.130*** (df = 3; 56)	21.208*** (df = 4; 55)
Note:				*p<0.1; ** p<0.05; ***p<0.01			

As we can see, the regression model to predict ARKK's return under the Fama-French model includes all four factors, market excess return, HML, SMB and MOM, as independent variables, has an R-squared of 0.607, indicating that the four factors can account for 60.7% of ARKK's monthly return. The intercept here is -0.091, showing that if all the four factors equal to zero, the expected monthly return of ARKK is -0.091%, but this number rarely has a meaning since it is nearly impossible for all the four factors to be zero at the same time. The coefficient of excess market return and SMB is highly statistically significant at 1% level here, the coefficient of HML is statistically significant at 10% level and the coefficient of MOM is not statistically significant at all. So, if holding all other three factors fixed, a one percentage point increase in the market excess return is expected to be associated with a 0.017 percentage point increase in ARKK's monthly excess return, and a one percentage point increase in the SMB is expected to be associated with an increase in ARKK's monthly excess return of 0.017 percentage point. The model's F statistic is 21.208 with 1% significance level, clearly rejecting the null hypothesis that all the coefficients are equal to 0. If we plot the residuals from the variables included in the data set, we can find that the residuals are almost distributed normally, and they are relatively homoscedastic.

The coefficient of MOM is not significant at all here, showing that the momentum effect for ARKK is not as powerful or consistent as it is for other equities in the market. And the insignificance also suggests that ARKK's return is not so closely related to the tendency of

stocks that performed well in the past to continue performing well in the future.

As we can see, when comparing the 7 regression models based on the Fama-French four-factor model, the model which takes market excess return, SMB and HML as its independent variables has the highest adjusted R-squared, the lowest AIC, and its BIC is pretty low at the same time. Its F statistic is also significant at 1% level. As a result, we can clearly tell that the model includes market excess return, SMB and HML is a better one, although the R-squared of it is lower than the four-factor model. The slight improvement in R-squared cannot compensate for the loss in adjusted R-squared, AIC and BIC value, so the three-factor model is still a better choice when we want to predict ARKK's return. The three-factor model has an R-squared of 0.603, slightly lower than 0.607, suggesting that the model can explain 60.3% of the ARKK's excess return. The intercept there is -0.104 and still, it does not have a practical meaning. All of the coefficients of the three independent variables are statistically significant and if holding all the other two variables unchanged, a one percentage point increase in the market excess return now is still expected to be associated with an increase of 0.017 percentage point; a one percentage point increase in the SMB is also expected to be associated with a 0.017 percentage point increase; a one percentage point increase in the HML is expected to be associated with a 0.005 percentage point decrease in ARKK's return. The AIC value here is -100.352, smaller than that of the four-factor regression of -98.929, and its adjusted R-squared is 0.582, which is higher than that of 0.578.

4.3 ARKK under the Black-Litterman Model

	Dependent variable:						
	ARKK.RF						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mkt.RF	0.019***		0.019***	0.017***	0.017***	0.019***	0.017***
	(0.003)		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
SMB		0.028***		0.018***	0.015***		0.014***
		(0.006)		(0.005)	(0.005)		(0.005)
HML						-0.004	-0.004
						(0.003)	(0.003)
Lagged.ARKK.RF			0.396***		0.344***	0.379***	0.330***
			(0.087)		(0.082)	(0.087)	(0.082)
Constant	-0.107***	-0.090***	-0.073***	-0.103***	-0.075***	-0.075***	-0.077***
	(0.015)	(0.017)	(0.015)	(0.013)	(0.014)	(0.015)	(0.014)
Observations	60	60	59	60	59	59	59
R^2	0.488	0.245	0.623	0.580	0.682	0.639	0.695
Adjusted R^2	0.479	0.232	0.609	0.565	0.665	0.619	0.673
Akaike Inf. Crit.	-89.062	-65.802	-102.998	-98.961	-111.118	-103.553	-111.562
Bayesian Inf. Crit.	-82.779	-59.519	-94.688	-90.583	-100.731	-93.165	-99.097
Residual Std. Error	0.111 (df = 58)	0.135 (df = 58)	0.097 (df = 56)	0.102 (df = 57)	0.090 (df = 55)	0.096 (df = 55)	0.089 (df = 54)
F Statistic	55.194*** (df = 1; 58)	18.818*** (df = 1; 58)	46.222*** (df = 2; 56)	39.321*** (df = 2; 57)	39.357*** (df = 3; 55)	32.415*** (df = 3; 55)	30.778*** (df = 4; 54)
Note:			*p<0.1; ** p<0.05; ***p<0.01				

The graph above helps us to analyze the regressions to predict ARKK's return under the Black-Litterman model. The model includes all four factors, market excess return, HML, SMB and lagged ARKK's monthly excess return as our view, as independent variables. We can find that the model has an R-squared of 0.695, indicating that these four factors can account for 69.5% of ARKK's monthly return, which is higher than the first regression including MOM as the independent variable. The intercept here is -0.077, showing that if all the four factors equal to zero, the expected monthly return of ARKK is -0.077%, but this number rarely has a meaning in the real world since it is nearly impossible for all the four factors to be zero at the same time. The coefficient of excess market return, SMB and lagged ARKK's monthly excess return is highly statistically significant at 1 % level and the coefficient of HML is not statistically significant at all. If holding all other three factors fixed, a one percentage point increase in the market excess return is expected to be associated with a 0.017 percentage point increase in ARKK's monthly excess return; a one percentage point increase in the SMB is expected to be associated with an increase in ARKK's monthly excess return of 0.014 percentage point; and a one percentage point change in lagged one-month ARKK's monthly excess return is expected to be associated with an increase in ARKK's monthly excess return of 0.33 percentage point. The model's F statistic is 30.778 with 1% significance level, clearly rejecting the null hypothesis that all the coefficients are equal to 0. If we plot the residuals from the variables included in the data set, we can find that they are almost distributed normally, and are relatively homoscedastic.

When looking at the regression result, we can find that despite HML, all other variable's

coefficients are statistically significant. The aim of using HML factor in the Fama-French model is to explain the difference in the portfolio based on their value versus growth characteristics. It is not a significant independent variable when predicting ARKK's return as ARKK is actively managed and its strategy is to invest primarily in growth stocks whose book-to-market ratios are low. Since the design of adding HML factor into the Fama-French model is to explain the return of a wider range of stocks and portfolios, it may not be applicable to ARKK as it is a growth-oriented fund.

From the picture, we can see that among the 7 regression models, the four-factor model has the highest R-squared as well as the adjusted R-squared, and its AIC value is the lowest among all the models. But at the same time, we can find that despite the slightly higher AIC as well as slightly lower R-squared and adjusted R-squared, the three-factor model without HML has the lowest BIC value, and since the coefficient of HML is not significant at all in the four-factor model, we suggest that the three-factor model with market excess return, SMB and lagged ARKK monthly excess return as independent variables is the best regression model to predict ARKK's return under Black-Litterman model.

V. Conclusion

By comparing the two regressions which are either the best under Fama-French model or Black-Litterman model, we can find that the model under Black-Litterman model can explain the return of ARKK more accurately as it has a higher R-squared of 68.2% and a higher adjusted R-squared of 66.5%, suggesting us to add the lagged ARKK's return rather than MOM as the fourth independent variable and exclude HML. The lagged ARKK return can be more useful than MOM when predicting ARKK's future performance as it reflects the fund's recent performance, which can be an important driver of future returns. ARKK is actively managed, as mentioned before, the portfolio's composition changes frequently based on its updated investment strategy, making its recent return a stronger indicator to predict the portfolio's potential performance in the future. After analyzing the portfolio's performance in the recent three years, we can find that the portfolio is volatile, its performance is quite inconsistent, and this can be attributed to the changing macroeconomic conditions. Expectations of Fed rate hikes

and the associated change in 10-year Treasury yields have a disproportionately big impact on the valuations of hypergrowth equities held by ARKK as most of the value of these growth stocks is based on the terminal value and it is particularly sensitive to even tiny basis point changes in the risk-free rate and the discount rate.

We have found that the regression under Black-Litterman model performs better, but there do have some limitations. The data set has a rather small sample size, we take in total 60 monthly data of ARKK from the past 5 years. If we expand the data set, maybe we can have a regression model which can predict ARKK more accurately than the current two models. The predicting variables we used to predict ARKK's return can also be changed to make the model fit better. In both regression models, we collect market's return data from the French's Data Library, but it is possible for us to use NASDAQ's return as the market return to predict ARKK's return as it includes Nasdaq-listed equities of technology and growth-oriented enterprises and most of ARKK's holdings fall in the technology and growth-oriented stocks. The use of NASDAQ return can more accurately represent the possible move in technology stocks and can exclude the effect of the return on other non-tech stocks when we try to predict ARKK's return.

Still, we can say that the use of market excess return, SMB and lagged one-month ARKK's return can help us to predict ARKK's future move accurately and despite the use of the regression model, investors who want to obtain the expected return from ARKK still also need to gain an accurate understanding of the overall economy.

Works Cited

ARKK. Yahoo Finance. https://finance.yahoo.com/quote/ARKK?p=ARKK&.tsrc=fin-srch.

Benali, Mimoun, et al. "Pricing Ability of Carhart Four-Factor and Fama–French Three-Factor Models: Empirical Evidence from Morocco." *International Journal of Financial Studies*, vol. 11, no. 1, Mar. 2023, p. 20. EBSCOhost, https://doi.org/10.3390/ijfs11010020.

Carhart, Mark M. "On persistence in mutual fund performance." *Journal of finance*, vol. 52, no.1, 1997, pp. 57-82.

Fama, Eugene F., and Kenneth R. French. "The cross‐section of expected stock returns." *Journal of Finance*, vol. 47, no. 2, 1992, pp. 427-465.

Fama, Eugene F., and Kenneth R. French. "Multifactor Explanations of Asset Pricing Anomalies." *The Journal of Finance*, vol. 51, no. 1, 1996, pp. 55–84., https://doi.org/10.1111/j.1540-6261.1996.tb05202.x.

He, Guangliang, and Robert Litterman. The intuition behind black-litterman model portfolios. http://dx.doi.org/10.2139/ssrn.334304

Ji, Ziyang, et al. "Empirical Research on the Fama-French Three-Factor Model and a Sentiment-Related Four-Factor Model in the Chinese Blockchain Industry." *Sustainability (Switzerland)*, vol. 12, no. 12, June 2020. EBSCOhost, https://doi.org/10.3390/su12125170.

Kenneth R. French's Data Library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Li, Jia-long, et al. "Model contest and portfolio performance: Black-Litterman versus factor models." *2013 International Conference on Management Science and Engineering 20th Annual Conference Proceedings*, Harbin, China, 2013, pp. 507-516, doi: 10.1109/ICMSE.2013.6586329.

Markowitz, Harry M.. *Portfolio Selection : Efficient Diversification of Investments*, Yale University Press, 1971. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/nyulibrary-ebooks/detail.action?docID=3421285.

NASDAQ Composite. Yahoo Finance.

https://finance.yahoo.com/quote/%5EIXIC?p=^IXIC&.tsrc=fin-srch.

Sharpe, W.F. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *Journal of Finance*, vol. 19, 1964, pp. 425-442.

Wen, Yiming, and Zongcheng Yin. “Study on Excess Stock Returns of Pharmaceutical and Biological Listed Companies Based on the Fama-French Three-Factor Model.” *China-Arab States Science and Technology Forum*, no. 9, 2022, p. 87. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=edscqv&AN=edscqv.7108050641&site=eds-live.